(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号
特開2003－44080
(P2003－44080A)
(43)公開日　平成15年２月14日(2003.2.14)

(51)Int.Cl.'	識別記号	FI		テーマコード'(参考)
G10L 15/06		G06K 9/00	S	5B064
G06K 9/00		G10L 3/00	521C	5D015
G10L 15/00			521J	
15/14			535Z	
15/20			551H	

審査請求　未請求　請求項の数22　ＯＬ　（全 25 頁）　最終頁に続く

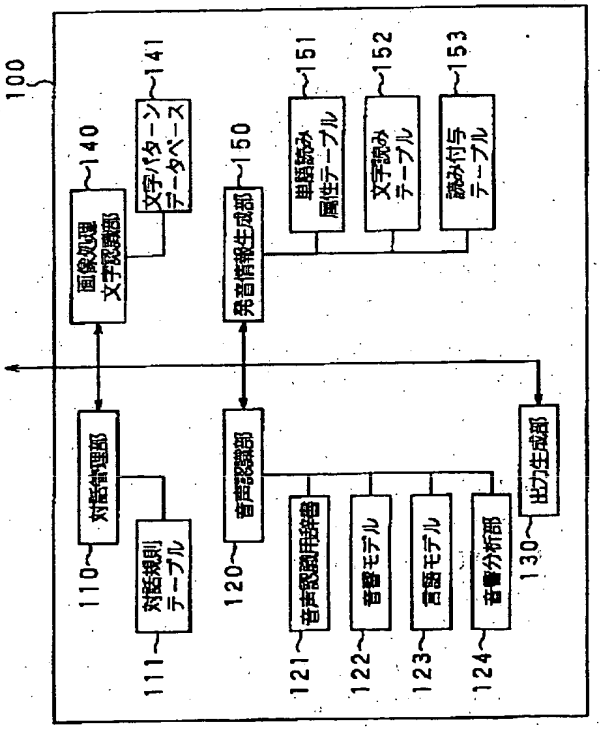
(21)出願番号　特願2002－130905(P2002－130905)
(22)出願日　平成14年５月２日(2002.5.2)
(31)優先権主張番号　特願2001－135423(P2001－135423)
(32)優先日　平成13年５月２日(2001.5.2)
(33)優先権主張国　日本（ＪＰ）

(71)出願人　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号
(72)発明者　廣江　厚夫
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(72)発明者　南野　活樹
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(74)代理人　100067736
弁理士　小池　晃　（外２名）

最終頁に続く

(54)【発明の名称】　ロボット装置、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体

(57)【要約】
【課題】　未登録の単語を新規単語として認識用辞書に登録する。
【解決手段】　ＣＣＤカメラ２０において撮像された画像の文字認識の結果から推定される複数の文字と、これら各文字から推定される複数の読み仮名と、各読み仮名に対応する読み方とを発音情報生成部１５０において生成し、ここで得られた複数の読み方とマイク２３において取得したユーザからの発声とをマッチングすることによって、生成された複数候補の中から１つの読み仮名及び発音のしかた（読み方）を特定する。



1

【特許請求の範囲】

【請求項１】　内部状態に応じて自律的に動作するロボット装置において、

単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、

単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、

被写体を撮像する撮像手段と、

上記撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、

周囲の音を取得する集音手段と、

上記集音手段において取得された音から音声を認識する音声認識手段と、

上記画像認識手段において抽出された上記所定パターンから推定される複数通りの表音文字を上記単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、

上記発音情報生成手段において生成された各音声波形と上記音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を上記画像認識手段において抽出されたパターン認識結果に対応する発音のしかたであるとして上記音声認識用辞書に新規に記憶する記憶制御手段とを備えることを特徴とするロボット装置。

【請求項２】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項１記載のロボット装置。

【請求項３】　上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時的に記憶する一時記憶手段を備えることを特徴とする請求項２記載のロボット装置。

【請求項４】　単語と該単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、上記記憶制御手段は、新規に記憶する文字と該文字の発音のしかたとともに上記単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求項２記載のロボット装置。

【請求項５】　上記音声認識手段において認識された音声に対する応答を生成する対話管理手段を備え、上記対話管理手段は、上記単語属性を音声に対する応答規則で使用することを特徴とする請求項４記載のロボット装置。

【請求項６】　上記音声認識手段は、隠れマルコフモデル法に基づいて音声を認識することを特徴とする請求項２記載のロボット装置。

【請求項７】　単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、

単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、

2

被写体を撮像する撮像手段と、

上記撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、

周囲の音を取得する集音手段と、

上記集音手段において取得された音から音声を認識する音声認識手段と、

上記画像認識手段において抽出された上記所定パターンの画像から推定される複数通りの表音文字を上記単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、

上記発音情報生成手段において生成された各音声波形と上記音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして上記音声認識用辞書に新規に記憶する記憶制御手段とを備えることを特徴とする文字認識装置。

【請求項８】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項７記載の文字認識装置。

【請求項９】　上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時的に記憶する一時記憶手段を備えることを特徴とする請求項７記載の文字認識装置。

【請求項１０】　単語と該単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、上記記憶制御手段は、新規に記憶する文字と該文字の発音のしかたとともに上記単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求項７記載の文字認識装置。

【請求項１１】　上記音声認識手段において認識された音声に対する応答を生成する対話管理手段を備え、上記対話管理手段は、上記単語属性を音声に対する応答規則で使用することを特徴とする請求項１０記載の文字認識装置。

【請求項１２】　上記音声認識手段は、隠れマルコフモデル法に基づいて音声を認識することを特徴とする請求項７記載の文字認識装置。

【請求項１３】　被写体を撮像する撮像工程と、

上記撮像工程において撮像された画像から所定パターンの画像を抽出する画像認識工程と、

周囲の音を取得する集音工程と、

上記集音工程において取得された音から音声を認識する音声認識工程と、

上記画像認識工程において抽出された所定パターンの画像から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、

3

上記発音情報生成工程において生成された各音声波形と上記音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備えることを特徴とする文字認識方法。

【請求項１４】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項１３記載の文字認識方法。

【請求項１５】　上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時記憶手段に記憶する工程を備えることを特徴とする請求項１４記載の文字認識方法。

【請求項１６】　上記記憶制御工程では、新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求項１４記載の文字認識方法。

【請求項１７】　上記音声認識工程において認識された音声に対する応答を生成する対話管理工程を備え、上記対話管理工程では、上記単語属性が音声に対する応答規則で使用されることを特徴とする請求項１６記載の文字認識方法。

【請求項１８】　上記音声認識工程では、隠れマルコフモデル法に基づいて音声認識されることを特徴とする請求項１４記載の文字認識方法。

【請求項１９】　内部状態に応じて自律的に動作するロボット装置の制御プログラムにおいて、
被写体を撮像する撮像処理と、
上記撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、
周囲の音を取得する集音処理と、
上記集音処理によって取得された音から音声を認識する音声認識処理と、
上記画像認識処理によって抽出された所定パターンの画像から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、
上記発音情報生成処理によって生成された各音声波形と上記音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させることを特徴とする制御プログラム。

【請求項２０】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項１９記載の制御プログラム。

4

【請求項２１】　被写体を撮像する撮像処理と、
上記撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、
周囲の音を取得する集音処理と、
上記集音処理によって取得された音から音声を認識する音声認識処理と、
上記画像認識処理によって抽出された所定パターンの画像から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、
上記発音情報生成処理によって生成された各音声波形と上記音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させるための制御プログラムが記録された記録媒体。

【請求項２２】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項２１記載の記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、内部状態に応じて自律的に動作するロボット装置、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体に関し、特に、撮像した画像から所定パターンの画像を認識し、この画像とともに取得した音声をこの認識画像と対応付けて新規に登録するロボット装置、並びに、撮像された所定パターンの画像とともに取得した音声をこの認識画像と対応付けて新規に登録する文字認識装置及び文字認識方法、並びに、取得した画像から所定パターンの画像を認識し、この画像とともに取得した音声をこの認識画像と対応付けて新規に登録する処理を実行させる制御プログラム及びこの制御プログラムが記録された記録媒体に関する。

【０００２】

【従来の技術】電気的又は磁気的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット」という。我が国においてロボットが普及し始めたのは、１９６０年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータや搬送ロボット等の産業用ロボット（Industrial Robot）であった。

【０００３】最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットの開発が進められている。このような実用ロボットは、産業用ロボットとは異なり、人間の生活環境の様々な局面におい

て、個々に個性の相違した人間、又は様々な環境への適応方法を自ら学習する能力を備えている。例えば、犬、猫のように４足歩行の動物の身体メカニズムやその動作を模した「ペット型」ロボット、或いは、２足直立歩行を行う動物の身体メカニズムや動作をモデルにしてデザインされた「人間型」又は「人間形」ロボット（Humanoid Robot）等の脚式移動ロボットは、既に実用化されつつある。これらの脚式移動ロボットは、動物や人間の容姿にできるだけ近い外観形状を有し、産業用ロボットと比較して動物や人間の動作に近い動作を行うことができ、更にエンターテインメント性を重視した様々な動作を行うことができるため、エンターテインメントロボットと呼称される場合もある。

【０００４】脚式移動ロボットの中には、「目」に相当する小型カメラや、「耳」に相当する集音マイク等を備えているものもある。この場合、脚式移動ロボットは、取得した画像に対して画像処理を施すことによって、画像情報として入力した周囲の環境を認識したり、入力した周囲の音から「言語」を認識したりできる。

【０００５】特に、外部から取得した音声を認識して文字に変換したり、音声を認識して応答したりする手法は、脚式移動ロボット以外にもパーソナルコンピュータや、その他の電子機器に音声認識装置として適用されている。

【０００６】従来の音声認識の手法では、単語の発音と表記とが対応付けされて記憶された音声認識用の辞書（以下、認識用辞書と記す。）を用いて音声認識している。そのため、認識用辞書に登録されていない単語に関しては認識できないという欠点があった。更に、「文」のような連続した単語の発音を認識する場合には、認識用辞書に登録されている単語の組み合わせでなくてはならない。つまり、認証用辞書に登録されていない単語が含まれる場合、誤認識されるか、認識できない。

【０００７】「北品川」という単語を例にとると、「北品川」が認証用辞書に登録されていなければ、「北品川」及び「北品川」を含む発音、例えば、「北品川は、どこですか。」という単語の連続からなる音声は、認識できないか、「北品川」の部分が誤認識される。そこで、認識用辞書に登録されていない単語を認識できるようにするためには、未登録の単語を新たに追加登録することが必要になる。

【０００８】音声認識装置が音声認識を可能とするために備える認識用辞書とは、他の単語と区別するための識別子としての「単語シンボル」と、その単語の発音情報を表す「ＰＬＵ列」とが対応付けられたものである。ＰＬＵ（Phonone-like unit）とは、音響的及び音韻的単位となるものである。発音された音声は、ＰＬＵの組み合わせ（ＰＬＵ列）として必ず表現することができる。

【０００９】したがって、認識用辞書に単語を登録する場合は、単語シンボルとこれに対応するＰＬＵ列とを追加すればよい。ただし、単語シンボルとＰＬＵ列とを追加できる場合というのは、「北品川」や「ｋｉｔａｓｈｉｎａｇａｗａ」という表記を、例えば、キーボード等のような入力手段を用いて直接入力できる場合に限られる。

【００１０】そのため、ロボット装置のようにキーボードのような入力手段を備えていない場合には、音声として取得した単語の発音を音声認識して未知単語のＰＬＵ列を得る方法もある。この場合、ガーベージモデル（garbage model）を適用して認識している。ガーベージモデルとは、図２０（ａ）及び図２０（ｂ）に示すように、音声を発音の基本的な単位となる「音素」の組み合わせとして表した、また、単語の読み方の基本的な単位となる「かな」の組み合わせとして表した（ただし、日本語の場合。）モデルである。

【００１１】従来の音声認識装置では、ガーベージモデルを適用することによって、音声による認識結果が得て、この認識結果に単語シンボルを当てはめて、これらを対応させて新規単語として認識用辞書に登録している。

【００１２】ただし、ここで「音素」と「ＰＬＵ」とは、ほぼ同義の単語として使用しており、「ＰＬＵ列」は、複数の「ＰＬＵ」が接続されることで構成された単語の発音を表記したものである。

【００１３】

【発明が解決しようとする課題】ところが、ガーベージモデルを適用した従来の音声認識の手法では、同じ単語であってもユーザ毎に発声のしかたに微妙な違いがあることや、弱い音素（例えば、語頭の／ｓ／等）は、必然的に認識されにくくなることや、周囲の雑音の影響による音素の変化や、音声区間検出の失敗等が原因となって、認識精度が悪くなるという欠点があった。

【００１４】特に、ロボット装置に音声認識装置を適用した場合、音声認識装置側の音声取得用のマイクとユーザ（音声源）との距離が離れている状況下で使用されることが多いため、誤認識の頻度が高くなる。

【００１５】具体的に、例えば、「きたしながわ」を認識させる場合について示すと、認識結果は、「ｈｉ　ｔｏ　ｔｓｕ　ｎａ　ｎｏ　ｇａ」や「ｉ　ｔａｓ　ｎａ　ｇａ：」のように「きたしながわ」と類似しているが、同一ではないＰＬＵ列として認識されることがある。このような方法で単語登録された辞書を用いて音声認識を行うと、認識精度の低下、また誤認識による表示誤り等の問題が発生する。つまり、新規登録語には、不正確なＰＬＵ列が付与されていることになるため、この単語を認識する際の精度が低下するという問題点があった。

【００１６】登録した人とは別の人が同じ単語を発音した場合、仮に「きたしながわ」が認識用辞書に登録されていたとしても、ユーザ毎の発音の癖から「きたしなが

わ」という単語を含む発音が認識されないこともあった。

【００１７】また、音声認識の結果を文字に変換して表示する場合、新規登録語には、表示に関する情報が与えられていないため、誤った文字が表示されることがある。ユーザが「きたしながわ」を音声で登録した後、音声認識装置に対して「北品川に行きたい。」と発声した場合、音声認識装置には「きたしながわ」が正しく認識されたとしても、表示は「ｈｉｔｏｔｓｕｎａｎｏｇａに行きたい」や「『ひとつなのが』に行きたい」になることがある。また、音声認識装置が認識結果のＰＬＵ列を音声合成で反復する場合も、合成された新規登録語のＰＬＵ列の部分だけが不自然な繋がりとして発声されるという不都合も生じる。

【００１８】更に、このようにガーベージモデルによって登録された新規登録語は、品詞や意味等の単語の属性に関する情報を登録することができない。例えば、「北品川」を登録したとしても、この単語が名詞であるか地名であるかを表す情報を登録することができない。そのため、仮に、例えば、対話用の文法や認識用の言語モデル等に「＜地名を表す語＞＋は＋どこ＋です＋か」のような特定表現のための文法規則が予め記録されていたとしても、新規登録語には適用できないという問題点があった。登録時に単語の属性についても音声で入力することができるが、ユーザが単語の属性を知っている必要があった。また、単語の登録操作に加えて属性を入力することはユーザにとって煩わしい。

【００１９】そこで本発明は、このような従来の実情に鑑みて提案されたものであり、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として認識することによって、未登録の単語を新規単語として認識用辞書に登録でき、更に登録された新規単語を精度よく認識できるロボット装置、並びに、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として認識することによって、未登録の単語を新規単語として認識用辞書に登録でき、登録された新規単語を精度よく認識できる文字認識装置、及び、提示された文字を撮像し、撮像された画像から文字を認識し、提示とともに発音された音声を取得して認識された文字の発音として認識することによって、認識用辞書に新規単語として登録する文字認識方法、並びに、撮像した画像から文字を認識し取得した音声をこの文字の発音として新規に登録する処理を実行させる制御プログラム及びこの制御プログラムが記録された記録媒体を提供することを目的とする。

【００２０】

【課題を解決するための手段】上述した目的を達成するために、本発明に係るロボット装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【００２１】このようなロボット装置は、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した所定パターンの画像に対応する発音のしかたであるとして音声認識用辞書に新規に記憶する。

【００２２】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。

【００２３】また、本発明に係る文字認識装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から文所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【００２４】このような文字認識装置は、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音

のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する。

【００２５】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。

【００２６】また、本発明に係る文字認識方法は、被写体を撮像する撮像工程と、撮像工程において撮像された画像から所定パターンの画像を抽出する画像認識工程と、周囲の音を取得する集音工程と、集音工程において取得された音から音声を認識する音声認識工程と、画像認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備える。

【００２７】このような文字認識方法によれば、画像認識工程において抽出された所定パターンの画像から推定される複数通りの表音文字が単語表音テーブルに基づいて付与され、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形が生成され、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とが比較され、最も近い音声波形が抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶される。

【００２８】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。

【００２９】更に、本発明に係る制御プログラムは、被写体を撮像する撮像処理と、撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、周囲の音を取得する集音処理と、集音処理によって取得された音から音声を認識する音声認識処理と、画像認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、発音情報生成処理によって生成された各音声波形と音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させる。

【００３０】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。また、上述の制御プログラムを記録媒体に記録して提供する。

【００３１】

【発明の実施の形態】本発明の一構成例として示すロボット装置は、内部状態に応じて自律動作するロボット装置である。このロボット装置は、少なくとも上肢と体幹部と下肢とを備え、上肢及び下肢、又は下肢のみを移動手段とする脚式移動ロボットである。脚式移動ロボットには、４足歩行の動物の身体メカニズムやその動きを模倣したペット型ロボットや、下肢のみを移動手段として使用する２足歩行の動物の身体メカニズムやその動きを模倣したロボット装置があるが、本実施の形態として示すロボット装置は、４足歩行タイプの脚式移動ロボットである。

【００３２】このロボット装置は、住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットであり、内部状態（怒り、悲しみ、喜び、楽しみ等）に応じて行動できるほか、４足歩行の動物が行う基本的な動作を表出できるエンターテインメントロボットである。

【００３３】このロボット装置は、特に「犬」を模した形体であり、頭部、胴体部、上肢部、下肢部、尻尾部等を有している。各部の連結部分及び関節に相当する部位には、運動の自由度に応じた数のアクチュエータ及びポテンショメータが備えられており、制御部の制御によって目標とする動作を表出できる。

【００３４】このロボット装置は、周囲の状況を画像データとして取得するための撮像部、周囲の音声を取得するマイク部、外部から受ける作用を検出するための各種センサ等を備えている。撮像部としては、小型のＣＣＤ（Charge-Coupled Device）カメラを使用する。

【００３５】本実施の形態として示すロボット装置は、画像認識装置及び音声認識装置を備えており、ＣＣＤカメラにおいて撮像された画像から所定パターンの画像を抽出し、抽出された所定パターンの画像から推定される複数通りの読み仮名を付与し、付与された複数通りの読み仮名のそれぞれに相当する音声波形を生成する。ここでの画像の所定パターンとしては、文字（文字列）、物体の形状、輪郭、柄、物体そのものの画像等があげられる。そして、この音声波形とマイク部において取得した音声の音声波形とを比較し、最も近い音声波形を抽出した所定パターンの画像に対応する発音のしかた（読み方）であるとして音声認識用辞書に新規に記憶することができるロボット装置である。

【００３６】以下、本発明の一構成例として示すロボット装置について、図面を参照して説明する。以下の説明では、取得した画像から認識される所定パターンが文字（文字列）である場合について詳細に説明する。

【００３７】本実施の形態では、ロボット装置１は、図

１に示すように、「犬」を模した形状のいわゆるペット型ロボットである。ロボット装置１は、胴体部ユニット２の前後左右に脚部ユニット３Ａ、３Ｂ、３Ｃ、３Ｄが連結され、胴体部ユニット２の前端部に頭部ユニット４が連結され、後端部に尻尾部ユニット５が連結されて構成されている。

【００３８】胴体部ユニット２には、図２に示すように、ＣＰＵ（Central Processing Unit）１０、ＤＲＡＭ（Dynamic Random Access Memory）１１、フラッシュＲＯＭ（Read Only Memory）１２、ＰＣ（Personal Computer）カードインターフェイス回路１３及び信号処理回路１４が内部バス１５を介して相互に接続されることにより形成されたコントロール部１６と、このロボット装置１の動力源としてのバッテリ１７とが収納されている。また、胴体部ユニット２には、ロボット装置１の向きや動きの加速度を検出するための角速度センサ１８及び加速度センサ１９が収納されている。

【００３９】頭部ユニット４には、外部の状況を撮像するためのＣＣＤ（Charge Coupled Device）カメラ２０と、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ２１と、前方に位置する物体までの距離を測定するための距離センサ２２と、外部音を集音するためのマイク２３と、鳴き声等の音声を出力するためのスピーカ２４と、ロボット装置１の「目」に相当するＬＥＤ（Light Emitting Diode）（図示せず）等が所定位置にそれぞれ配置されている。ＣＣＤカメラ２０は、頭部ユニット４の向く方向にある被写体を所定の画角で撮像することができる。

【００４０】各脚部ユニット３Ａ～３Ｄの関節部分、各脚部ユニット３Ａ～３Ｄと胴体部ユニット２との連結部分、頭部ユニット４と胴体部ユニット２との連結部分、尻尾部ユニット５と尻尾５Ａとの連結部分には、自由度数分のアクチュエータ２５$_1$～２５$_n$及びポテンショメータ２６$_1$～２６$_n$がそれぞれ配設されている。アクチュエータ２５$_1$～２５$_n$は、例えば、サーボモータを構成として有している。サーボモータの駆動により、脚部ユニット３Ａ～３Ｄが制御されて目標の姿勢、或いは動作に遷移する。

【００４１】これら角速度センサ１８、加速度センサ１９、タッチセンサ２１、距離センサ２２、マイク２３、スピーカ２４及び各ポテンショメータ２６$_1$～２６$_n$等の各種センサ並びにＬＥＤ及び各アクチュエータ２５$_1$～２５$_n$は、それぞれ対応するハブ２７$_1$～２７$_n$を介してコントロール部１６の信号処理回路１４と接続され、ＣＣＤカメラ２０及びバッテリ１７は、それぞれ信号処理回路１４と直接接続されている。

【００４２】信号処理回路１４は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス１５を介して

ＤＲＡＭ１１内の所定位置に順次格納する。また信号処理回路１４は、これとともにバッテリ１７から供給されるバッテリ残量を表すバッテリ残量データを順次取り込み、これをＤＲＡＭ１１内の所定位置に格納する。

【００４３】このようにしてＤＲＡＭ１１に格納された各センサデータ、画像データ、音声データ及びバッテリ残量データは、ＣＰＵ１０が当該ロボット装置１の動作制御を行う際に使用される。

【００４４】ＣＰＵ１０は、ロボット装置１の電源が投入された初期時において、フラッシュＲＯＭ１２に格納された制御プログラムを読み出して、ＤＲＡＭ１１に格納する。又は、ＣＰＵ１０は、図１に図示しない胴体部ユニット２のＰＣカードスロットに装着された半導体メモリ装置、例えば、いわゆるメモリカード２８に格納された制御プログラムをＰＣカードインターフェイス回路１３を介して読み出してＤＲＡＭ１１に格納する。

【００４５】ＣＰＵ１０は、上述のように信号処理回路１４よりＤＲＡＭ１１に順次格納される各センサデータ、画像データ、音声データ、及びバッテリ残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無を判断している。

【００４６】更に、ＣＰＵ１０は、この判断結果とＤＲＡＭ１１に格納した制御プログラムとに基づく行動を決定する。ＣＰＵ１０は、当該決定結果に基づいてアクチュエータ２５$_1$～２５$_n$の中から必要とするアクチュエータを駆動することによって、例えば、頭部ユニット４を上下左右に動かしたり、尻尾部ユニット５の尻尾を動かしたり、各脚部ユニット３Ａ乃至３Ｄを駆動して歩行させたりする。また、ＣＰＵ１０は、必要に応じて音声データを生成し、信号処理回路１４を介してスピーカ２４に供給する。また、ＣＰＵ１０は、上述のＬＥＤの点灯・消灯を指示する信号を生成し、ＬＥＤを点灯したり消灯したりする。

【００４７】また、ＣＰＵ１０は、上述のようにロボットを自律的に制御するほかに、後述する対話管理部１１０等からの要求に応じてロボットを動作させる。

【００４８】これらの基本的な構成によって、ロボット装置１は、自己及び周囲の状況や、使用者からの指示及び働きかけに応じて自律的に行動する。

【００４９】更に、ロボット装置１は、認識した発音と認識した文字との対応を新規登録語として音声認識用辞書に登録するための構成として、胴体部ユニット２のコントロール部１６に、画像音声認識部１００を備えている。画像音声認識部１００は、図３に示すように、対話管理部１１０と、音声認識部１２０と、出力生成部１３０と、画像処理文字認識部１４０と、発音情報生成部１５０とを有している。音声認識用辞書とは、図４に示すように、他の単語と区別するための識別子としての「単語シンボル」と、この単語に対応する発音情報を表す「ＰＬＵ列」とを記録したテーブルである。この辞書を

参照することによって、単語の発音のしかた（読み方）、又は、発音に対応する単語の表記が抽出できる。

【００５０】具体的に、対話管理部１１０は、マイク２３から入力したユーザの発話、対話履歴等から入力した音声に対する応答を生成する。対話管理部１１０は、対話規則テーブル１１１に記憶された種々の対話規則に基づいて、入力した音声に対する応答パターンを生成する。

【００５１】音声認識部１２０は、ユーザの発話を対話管理部１１０で処理できる形式、例えば、テキスト形式、構文解析、対話用フレーム等に変換する。音声認識部１２０は、具体的には、音声認識用辞書１２１、音響モデル１２２、言語モデル１２３、音響分析部１２４等からなる。音響分析部１２４では、認識に必要な特徴量の抽出が微少時間間隔で行われる。例えば、得られた音声信号のエネルギ、零交差数、ピッチ、周波数特性、及びこれらの変化量等が抽出される。周波数分析には、線形予測分析（ＬＰＣ）、高速フーリエ変換（ＦＦＴ）、バンドパスフィルタ（ＢＰＦ）等が用いられる。

【００５２】音声認識部１２０は、音響モデル１２２と言語モデル１２３とを用いて、音響分析部１２４で生成された特徴量系列に対応する単語系列を決定する。認識手法としては、例えば、隠れマルコフモデル（Hidden Markov Model：以下、ＨＭＭと記す。）等が用いられる。

【００５３】ＨＭＭとは、状態遷移確率と確率密度関数とをもつ状態遷移モデルであり、状態を遷移しながら特徴量系列を出力する確率値を累積して尤度を決定する。その尤度の値を「スコア」として音声認識用辞書に記憶されている単語の発音のしかたと後述する画像処理文字認識部において認識された文字に対して付与される発音のしかたとのマッチングに使用する手法である。ＨＭＭの遷移確率及び確率密度関数等は、学習用データに基づく学習過程を通じて、予め学習して用意される値である。

【００５４】音響モデルは、音素（ＰＬＵ）、音節、単語、フレーズ、文等、それぞれの単位毎に用意することができる。例えば、日本語の仮名『あ』・『い』・『う』・『え』・『お』・『か』・『き』…『ん』を単位とする音響モデルを用いた場合、これらを組み合わせて接続することによって、『はい』、『いいえ』、『おはよう』、『いまなんじですか』等の言葉が構成できる。音素とは、単語の発音情報を表すものであり、音響的及び音韻的単位である。本明細書では、音素とＰＬＵ（Phonone-like unit）とを区別しないで使用している。発音された音声は、音素（ＰＬＵ）の組み合わせ（ＰＬＵ列）として必ず表現することができる。

【００５５】ＨＭＭによれば、このように構成された言葉とマイク２３において取得した音声の特徴量系列との類似度をスコアとして計算することができる。音響モデルから「言葉」を構成するための情報として、言語モデル１２３と音声認識用辞書１２１とが利用される。音声認識用辞書１２１とは、認識対象となる各単語を構成するための音響モデル（ここでは、仮名の一文字『あ』、『い』、・・・等を示す。）の接続のしかたを対応テーブルとして示した辞書であり、言語モデル１２３とは、単語と単語との接続のしかたの規則を示したものである。

【００５６】以下に示す例では、「単語」とは、認識処理の上で発音する際に、１つの纏まりとして扱う方が都合がよい単位のことを示しており、言語学的な単語とは必ずしも一致しない。例えば、以下の例では「北品川」を一単語として扱う場合があるが、これを「北」「品川」という２単語として扱っても構わない。更に、「北品川駅」や「北品川駅はどこですか」を発音する上での一単語として扱うこともできる。

【００５７】また、本明細書では、「読み仮名」とは、漢字、英単語の読み方を表記したひらがな又はカタカナの意として用い、「発音のしかた」とは、読み仮名の実際の発音をローマ字、又はローマ字と記号とを使用して表記したものであり、言語学的における「音素記号」に相当する。

【００５８】例えば、『～時から、～時まで』という文を扱う場合について考える。この場合、まず、『０（ぜろ）』『１（いち）』・・・『２４（にじゅうよん）』という単語と、『時（じ）』・『から』・『まで』という言葉のそれぞれに関して、音響モデル１２２を参照することによって、単語の接続のしかたが決定される。

【００５９】次に、『（数字を表す単語）』、『時』、『から』、『（数字を表す単語）』、『時』、『まで』という各単語を言語モデル１２３を参照することによって、文を構成するための各単語の接続のしかたが決定される。

【００６０】この音声認識用辞書１２１と言語モデル１２３とを用いてＨＭＭを適用することによって、『１時から２時まで』や『２時から５時まで』等の文と入力される特徴量系列との類似度がスコアとして計算できる。その中で最も高いスコアを有する単語系列からなる文を音声認識結果として出力する。

【００６１】音声認識処理におけるスコアの計算は、音響モデル１２２によって与えられる音響的なスコアと、言語モデル１２３によって与えられる言語的なスコアとを総合評価して行われる場合もある。

【００６２】言語的なスコアとは、例えば、連続するｎ個の単語間の遷移確率、又は連鎖確率に基づいて与えられるスコアである。遷移確率は、予め、大量のテキストから統計的に求められた値であり、ここでは、この遷移確率を「ｎグラム」と呼称する。

【００６３】なお、言語モデルは、文法やｎグラム中に単語を直接記述する以外にも、単語のクラス（単語をあ

る基準や属性にしたがって分類したもの）を記述する場合もある。

【００６４】例えば、地名を表す単語を集め、それに＜地名＞というクラス名称を与えた場合に「＜地名＞＋は＋どこ＋です＋か」という文法を記述したり、ｎグラム中に「＜地名＞＋は＋どこ」の遷移確率を用意しておくこともできる。この場合、ｎ＝３であり、正確には、遷移確率は、Ｐ（＜地名＞｜は、どこ｜）である。

【００６５】出力生成部１３０は、対話管理部１１０が生成した応答パターンを実際の動作に変換する。例えば、対話管理部１１０が「首を左右に振る＋『いいえ』と発声する」という応答パターンを生成した場合、出力生成部１３０は、これを受けて「首を左右に振る」に対応する動作パターンを生成しＣＰＵ１０に送るとともに、「いいえ」に対応する音声波形を生成しスピーカ２４から出力する。

【００６６】画像処理文字認識部１４０は、ＣＣＤカメラ２０で取り込んだ画像に含まれる文字列を文字パターンデータベース１４１に基づいて識別する。文字パターンデータベース１４１には、ひらがな、カタカナ、漢字、アルファベット、記号類、必要に応じて各国語の文字等の画像パターンが格納されている。画像処理文字識別部１４０は、ＣＣＤカメラ２０からの入力画像と文字パターンデータベース１４１に格納されている画像パターンとの間でマッチングを行い、入力画像に含まれている文字列を認識する。

【００６７】発音情報生成部１５０は、画像処理文字認識部１４０で認識された文字列に対応する発音情報、つまり文字列の読み仮名を生成し、更にその発音のしかた（読み方）を生成する。例えば、入力画像から「北品川」という文字列が認識された場合、「きたしながわ」という読み仮名を生成し、ＰＬＵ列で「ｋｉｔａｓｈｉｎａｇａｗａ」という発音のしかた（読み方）を生成する。

【００６８】単語読み属性テーブル１５１は、図４に示すように、単語（文字列）と読み仮名と属性の組を記述したテーブルである。属性とは、「地名」、「名前」、「動物」のように単語のもつ意味を示している。

【００６９】画像処理文字認識部１４０で認識された文字列がこのテーブルに含まれている場合は、このテーブルから読み仮名を抽出することで、読み仮名からその文字列の発音のしかた（読み方）を確定できる。単語読み属性テーブル１５１は、音声認識用辞書１２１とは独立に用意する。

【００７０】認識用辞書の語彙数には、認識速度や精度や処理上の都合で上限がある（例えば６万５５３６語）が、単語読み属性テーブル１５１にはそれらの制限とは関係なく単語を記述することができる。この単語読み属性テーブル１５１は、他の言語資源から流用することも可能である。例えば、仮名漢字変換プログラムや形態素解析プログラム等で使用されている辞書等を流用することもできる。

【００７１】文字読みテーブル１５２は、図６に示すように、文字と読み仮名との対応が記述されたテーブルである。記号やアルファベットや単漢字毎に読み仮名を記述しておく。使用可能な文字全てについて読み仮名を記述しておけば、任意の文字列に対して読み仮名から発音のしかた（読み方）を付与することができる。

【００７２】読み付与テーブル１５３は、２つのテーブルだけでは読み仮名が付与できない場合に読み仮名を付与するための規則や、読み仮名が特定できない場合に、これを特定するための規則が記述してある。例えば、音読み及び訓読みの統一、長音化に関する規則、連濁の規則、繰り返し記号に関する規則、英単語に読みを付与する規則がある。

【００７３】具体的には、長音化に関する規則とは、「・・・おう」「・・・えい」等を「・・・おー」「・・・えー」等に変換する規則である。この規則によって、例えば、「とうきょう」は、「とーきょー」に変換される。連濁の規則とは、例えば、「品川口」の読みを「しながわ（品川）」と「くち（口）」との結合から生成する場合に、「くち」を濁らせて「ぐち」にする規則である。また、繰り返し記号に関する規則とは、「々・ヽ・ヾ・ゝ・ゞ」等の繰り返し記号に対応して読み仮名を付ける規則である。更に、英単語に読み仮名を付与する規則とは、英単語の語末に“ｅ”がある場合は、“ｅ”自体は、発音しないかわりに前の母音を母音読みする等の規則である。例えば、“take”に「テーク」という読み仮名を付与する際に、“ａ”に対して「エー」という読み仮名を付与し、“ke”に対して、単に「ク」という読み仮名を付与する規則である。

【００７４】次に、認識用辞書に新規単語を登録する際の処理を、図７を用いて具体的に説明する。

【００７５】まず、ステップＳ１において、単語登録のための単語登録モードに移行する。単語登録モードへの移行は、例えば、ロボット装置１は、ユーザが発する「登録モード」や「言葉を覚えて」等の言葉をトリガとして単語登録モードに移行する。このほかに、操作ボタンを設け、この操作ボタンが押されたときに単語登録モードへ移行するようにしてもよい。

【００７６】ステップＳ２において、ロボット装置１は、ユーザに対して、登録したい単語の表記をロボット装置１のＣＣＤカメラ２０の前に提示する旨の指示及び／又は提示に加えてユーザが登録したい単語の読み方を発声する旨の指示を促す。ユーザに対する指示は、ロボット装置１が音声によって指示してもよいし、また、図示しない表示部に指示内容を表示する場合でもよい。ここでは、「北品川」という単語を例として説明する。ユーザによって提示される文字は、漢字でも仮名でもローマ字表記でもＰＬＵ列でも構わない。具体的には、ロボ

ット装置１は、「北品川」、「きたしながわ」、「キタシナガワ」、「ｋｉｔａｓｈｉｎａｇａｗａ」等の何れの表記も認識できる。

【００７７】ステップＳ３において、ロボット装置１は、文字提示のみであるか、文字提示とともに発話があったかを判断する。文字提示だけの場合は、ステップＳ４へ進み、文字提示とともに発話があった場合は、後述するステップＳ８へと進む。それ以外、すなわち、発声のみの場合は、従来と同様にガーベージモデルによる認識処理を行う。

【００７８】はじめに、文字提示のみの場合について説明する。文字提示のみの場合、ステップＳ４において、ロボット装置１における画像処理文字認識部１４０は、ＣＣＤカメラ２０において撮像された画像にどのような文字列が含まれているかを文字パターンデータベース１４１に基づいて、文字認識（ＯＣＲ：Optical Character Recognition）する。ここで、画像処理文字認識部１４０は、文字認識結果の候補が１つに絞り込めない場合、複数の候補を残す。例えば、「北品川」という文字に対して「比晶川」という認識結果が得られた場合は、「比晶川」も残す。

【００７９】続いて、ステップＳ５において、ロボット装置１における発音情報生成部１５０は、ステップＳ４での認識結果として得られた文字列に対して、文字列の発音のしかた（読み方）を生成する。発音を生成する際の詳細は、後述する。発音生成処理によって、文字列に対して発音のしかた（読み方）が付与される。認識された文字列が複数ある場合及び／又は１つの文字列に対して複数の発音のしかたが有り得る場合には、全ての発音パターンが適用される。

【００８０】ステップＳ６において、ロボット装置１は、上述のように生成された文字列に対する発音のしかた（読み方）が正しいか否か、又は、複数の読み方のうちどれを採用すべきかをユーザに確認する。発音のしかた（読み方）が一通りのみの場合は、「読み方は、○○で正しいですか。」のように質問する。ユーザが「正しい」や「はい」等の応答を返した場合は、ステップＳ７に進む。

【００８１】また、発音のしかた（読み方）が複数通りある場合は、それぞれについて「読み方は、○○ですか。」のように質問する。ユーザが「正しい」や「はい」等の応答を返した読み方を採用してステップＳ７に進む。

【００８２】ユーザから「いいえ」等の応答を受けた場合、すなわち、正しい読み方が存在しない場合、ステップＳ２若しくはステップＳ４の処理まで戻る。

【００８３】以上の処理によって、新規単語の読みを確定した後、ステップＳ７に進み、取得した文字列とこの文字列に対する発音のしかた（読み方）とを対応付けて新規単語として認識用辞書に登録する。新規単語を追加

する際、図４に示す単語シンボル欄には、提示された文字の認識結果を使用する。この文字列に対応するＰＬＵ列欄には、ステップＳ６において確定した発音のしかた（読み方）が記述される。新規単語を登録した後、登録モードを終了する。その後、更新された認識用辞書を音声認識に反映させるための処理、例えば、音声認識プログラムの再起動等を行う。

【００８４】一方、ステップＳ３において、ユーザが文字を提示するとともに表記した文字を発声した場合について説明する。文字提示ともに発話があった場合は、両者から得られる情報を協調的に使用することによってＰＬＵ列等の発音情報を精度よく生成することができる。

【００８５】具体的には、文字認識の結果から推定される複数の文字と、これら各文字から推定される複数の読み仮名と、各読み仮名に対応する発音のしかた（読み方）とを生成する。このようにして得られた複数の発音のしかた（読み方）とマイク２３において取得したユーザからの発声とをマッチングすることによって、上述のように生成された複数候補の中から１つの読み仮名及び発音のしかた（読み方）を特定する。

【００８６】文字提示とともに発話があった場合、ステップＳ８において、ロボット装置１における画像処理文字認識部１４０は、ＣＣＤカメラ２０において撮像された画像から文字認識する。ここで、画像処理文字認識部１４０は、文字認識結果の候補が１つに絞り込めない場合、複数の候補を残す。

【００８７】続いて、ステップＳ９において、ロボット装置１における発音情報生成部１５０は、ステップＳ８での認識結果として得られた文字列に対して、文字列の読み仮名を生成する。発音生成処理によって、文字列に対して発音のしかた（読み方）が付与される。認識された文字列が複数ある場合及び／又は１つの文字列に対して複数の読み方が可能な場合には、全ての発音パターンが適用される。

【００８８】次に、ステップＳ１０において、文字列と発音のしかた（読み方）とから、一時的に仮の認識用辞書を生成する。この辞書を以下、新規単語用認識用辞書と記す。例えば、ＣＣＤカメラ２０によって撮像された「北品川」という文字が画像処理文字認識部１４０において、「北品川」と「比晶川」の２通りに認識されたとする。音声情報生成部１５０は、「北品川」と「比晶川」に読み仮名を付与する。「北品川」には「きたしながわ」が付与され、「比晶川」には「ひしょうがわ」と「くらあきらがわ」の２通りが付与され、更に両者の発音のしかた（読み方）、すなわち、ＰＬＵ列が生成される。この場合の新規単語用認識用辞書を図８に示す。

【００８９】ステップＳ１１において、新規単語用認識用辞書を用いて、ユーザからの発声に対して音声認識を行う。ここでの音声認識は、連続音声認識ではなく、単語音声認識である。新規単語用認識用辞書が生成される

よりも前にユーザが発話している場合は、その発話を録音しておき、その録音音声に対して音声認識を行う。ステップＳ１１における音声認識とは、新規単語用認識用辞書に登録されている単語の中からユーザの発話と音響的に最も近い単語を探し出すことである。ただし、ステップＳ１１の処理では、単語シンボルが同一であっても、ＰＬＵ列が異なる場合は別の単語とみなす。

【００９０】図８では、ここに登録されている３単語（２つの「比晶川」は別単語とみなす）の中から、ユーザの発話である「きたしながわ」に最も近い単語を探し出すことである。結果として、単語シンボルとＰＬＵ列との組を１つに特定することができる。

【００９１】新規単語用認識用辞書の中から単語シンボルとＰＬＵ列との組が特定されたら、ステップＳ７において、これを正規の音声認識用辞書１２１に登録する。新規単語を登録した後、登録モードを終了する。その後、更新された認識用辞書を音声認識に反映させるための処理、例えば、音声認識プログラムの再起動等を行う。

【００９２】以上示した処理によって、ロボット装置１は、音声認識用辞書１２１に記憶されていない単語を新規単語として登録できる。

【００９３】上述したステップＳ５とステップＳ９での文字列の発音のしかた（読み方）の生成に関して、図９を用いて詳細に説明する。

【００９４】まず、ステップＳ２１において、画像処理文字認識部１４０によって認識された文字列が仮名文字だけで構成されているか否かを調べる。ただし、ここでの仮名文字とは、ひらがな・カタカナのほかに長音記号「ー」や繰り返し記号「々…」等も含む。文字列が仮名文字だけで構成されている場合は、ステップＳ２２において、認識された仮名文字をその文字列の読み方とする。このとき、長音化等の発音を若干修正する場合もある。

【００９５】一方、ステップＳ２１において、画像処理文字認識部１４０によって認識された文字列が仮名文字以外の文字を含んでいる場合、ステップＳ２３において、その文字列が単語読み属性テーブル１５１に含まれているか否かを判別する。

【００９６】文字列が単語読み属性テーブル１５１に含まれている場合は、そのテーブルから読み仮名を取得し、更に発音のしかた（読み方）を生成する（ステップＳ２４）。また、単語読み属性テーブル１５１に単語の属性が記述されている場合は、属性も同時に取得する。この属性の利用方法については、後述する。

【００９７】文字列が単語読み属性テーブル１５１に含まれていない場合、ステップＳ２５において、最長一致法・分割最小法、文字読みテーブル１５２に基づく読み付与、及び読み付与規則に基づく読み付与を組み合わせて読み仮名を取得する。

【００９８】最長一致法・分割数最小法とは、単語読み属性テーブル１５１に含まれる単語を複数組み合わせることで入力文字列と同じものが構成できないか試みる方法である。例えば、入力文字列が「北品川駅前」である場合、これが単語読み属性テーブル１５１に含まれていなくても「北品川」と「駅前」とが含まれていれば、これらの組み合わせから「北品川駅前」が構成できることから、結果として「きたしながわえきまえ」という読み方が取得できる。構成方法が複数通りある場合は、より長い単語が含まれる方を優先する（最長一致法）か、より少ない単語で構成できる方を優先する（分割数最小法）かして構成方法を選択する。

【００９９】また、文字読みテーブル１５２に基づく読み付与とは、文字列を文字毎に分割し、分割した文字毎に文字読みテーブル１５２から読み仮名を取得する方法である。漢字の場合、１つの漢字には複数の読み仮名が付与できるため、文字列全体としての読み仮名は、各漢字の読み仮名の組み合わせになる。そのため、例えば、「音読みと訓読とは混在しにくい」等の規則を用いて組み合わせの数を減らす方法である。

【０１００】続いて、ステップＳ２６において、上述の各方法で取得したそれぞれの読み仮名の候補に対してスコア又は信頼度を計算し、高いものを選択する。これにより、入力された文字列に読み仮名を付与できる。得られた読み仮名から発音のしかた（読み方）を生成する。

【０１０１】ステップＳ２２、ステップＳ２４、ステップＳ２６のそれぞれの工程を経たのち、最終的に、ステップＳ２７において、読み仮名に対する発音のしかた（読み方）を長音化や連濁化等の規則に基づいて修正する。

【０１０２】ここで、単語読み属性テーブル１５１について詳細に説明する。音声認識用辞書１２１に単語を新規登録しただけでは、言語モデル１２３に記録された単語間の接続規則を適用することはできない。例えば、「北品川」を音声認識用辞書１２１に追加登録したとしても、それだけでは「北品川」に関する文法や「北品川」と他の単語との連鎖確率等は、生成されない。したがって、新規登録語に言語モデルの接続規則を反映させる方法は、理想的には、文法を追加したり、テキストデータから連鎖確率を計算し直したりして、言語モデルを構成し直すことであるが、以下に示す簡易的な方法によって新規登録後に言語モデルを適用することができる。

【０１０３】まず、言語モデルに含まれていない単語に＜未知語＞というクラス名を付ける。言語モデルには＜未知語＞と他の単語との連鎖確率を記述しておく。新規登録語は、＜未知語＞とみなし、この新規登録語と他の単語との連鎖確率は、＜未知語＞と他の単語との連鎖確率から計算する。

【０１０４】クラスとは、単語をある基準や属性にしたがって分類したものである。例えば、意味にしたがって

分類し、それぞれを＜地名＞、＜姓＞、＜国名＞と命名したり、品詞にしたがって分類し、それぞれを＜名詞＞、＜動詞＞、＜形容詞＞と命名したりする。

【０１０５】言語モデルには、単語間の連鎖確率を記述するかわりにクラス間の連鎖確率やクラスと単語との連鎖確率を記述する。単語間の連鎖確率を求めるときは、単語がどのクラスに属すかを調べ、次に対応するクラスについての連鎖確率を求め、そこから単語間の連鎖確率を計算する。

【０１０６】新規登録語についても、どのクラスに属する単語であるかを登録時に推定することでクラスモデルが適用できる。

【０１０７】上述のようにすると未知語用モデルでは、新規登録語には、全て同一の値の連鎖確率が付される。それに対してクラスモデルでは、どのクラスに属するかによって異なる値になる。そのため一般的には、新規登録語についての言語的スコアは、クラスモデルを用いた方がより適切なスコアとなり、結果的に適切に認識される。

【０１０８】したがって、音声認識による単語登録において、従来、困難であったクラス名称が、容易に入力できる。すなわち、文字認識で得られた文字列（単語）が単語読み属性テーブル１５１に含まれている場合、このテーブルの属性欄からクラス名称を取得できる。なお、図５に示す例では、属性欄に属性を１つしか記述してないが、これを「＜地名＞、＜固有名詞＞、＜駅名＞」のように複数記述することもできる。この場合、例えば、＜地名＞というクラスが存在する場合は、＜地名＞、＜固有名詞＞、＜駅名＞の中から、クラス名称と一致する分類名、すなわち＜地名＞を採用する。

【０１０９】文字認識では、一文字ずつ認識するよりも、文字の連鎖に関する情報を含めて認識する方が精度が向上する場合がある。そこで、認識用辞書の「単語シンボル」欄や、単語読み属性テーブル１５１の「単語」欄等を文字の連鎖に関する情報として使用することによって、文字認識の精度を更に向上できる。

【０１１０】以上の説明では、取得画像における所定パターンの認識として文字認識の場合に関して説明したが、上述したように文字（文字列）のほか、物体の形状、輪郭、柄、物体そのものの画像を認識し対応する文字（文字列）を抽出し、抽出された文字から推定される複数通りの読み仮名を付与し、付与された複数通りの読み仮名のそれぞれに相当する音声波形を生成することもできる。この場合は、図１に示した基本的な構成に加えて、必要な構成が必要に応じて追加される。

【０１１１】このように、所定パターンとして文字列以外にも種々のケースに対応して発音のしかたをマスターできるようにすることにより、ロボット装置が外部から情報を得て学習していく様子を表現でき、エンターテインメント性が向上できる。

【０１１２】ところで、本実施の形態として示すロボット装置１は、内部状態に応じて自律的に行動できるロボット装置である。ロボット装置１における制御プログラムのソフトウェア構成は、図１０に示すようになる。この制御プログラムは、上述したように、予めフラッシュＲＯＭ１２に格納されており、ロボット装置１の電源投入初期時において読み出される。

【０１１３】図１０において、デバイス・ドライバ・レイヤ３０は、制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット３１から構成されている。この場合、各デバイス・ドライバは、ＣＣＤカメラ２０（図２）やタイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。

【０１１４】また、ロボティック・サーバ・オブジェクト３２は、デバイス・ドライバ・レイヤ３０の最下位層に位置し、例えば上述の各種センサやアクチュエータ２５$_1$～２５$_n$等のハードウェアにアクセスするためのインターフェイスを提供するソフトウェア群でなるバーチャル・ロボット３３と、電源の切換え等を管理するソフトウェア群でなるパワーマネージャ３４と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ３５と、ロボット装置１の機構を管理するソフトウェア群でなるデザインド・ロボット３６とから構成されている。

【０１１５】マネージャ・オブジェクト３７は、オブジェクト・マネージャ３８及びサービス・マネージャ３９から構成されている。オブジェクト・マネージャ３８は、ロボティック・サーバ・オブジェクト３２、ミドル・ウェア・レイヤ４０、及びアプリケーション・レイヤ４１に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ３９は、メモリカード２８（図２）に格納されたコネクションファイルに記述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

【０１１６】ミドル・ウェア・レイヤ４０は、ロボティック・サーバ・オブジェクト３２の上位層に位置し、画像処理や音声処理等のこのロボット装置１の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ４１は、ミドル・ウェア・レイヤ４０の上位層に位置し、当該ミドル・ウェア・レイヤ４０を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置１の行動を決定するためのソフトウェア群から構成されている。

【０１１７】なお、ミドル・ウェア・レイヤ４０及びアプリケーション・レイヤ４１の具体なソフトウェア構成をそれぞれ図１１に示す。

【０１１８】ミドル・ウェア・レイヤ４０は、図１１に

示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール５０～５８並びに入力セマンティクスコンバータモジュール５９等を有する認識系６０と、出力セマンティクスコンバータモジュール６８並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、ＬＥＤ点灯用及び音再生用の各信号処理モジュール６１～６７等を有する出力系６９とから構成されている。

【０１１９】認識系６０の各信号処理モジュール５０～５８は、ロボティック・サーバ・オブジェクト３２のバーチャル・ロボット３３によりＤＲＡＭ１１（図２）から読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール５９に与える。ここで、例えば、バーチャル・ロボット３３は、所定の通信規約によって、信号の授受或いは変換をする部分として構成されている。

【０１２０】入力セマンティクスコンバータモジュール５９は、これら各信号処理モジュール５０～５８から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」等の自己及び周囲の状況や、使用者からの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ４１に出力する。

【０１２１】アプリケーション・レイヤ４１は、図１２に示すように、行動モデルライブラリ７０、行動切換えモジュール７１、学習モジュール７２、感情モデル７３及び本能モデル７４の５つのモジュールから構成されている。

【０１２２】行動モデルライブラリ７０には、図１３に示すように、「バッテリ残量が少なくなった場合」、「転倒復帰する」、「障害物を回避する場合」、「感情を表現する場合」、「ボールを検出した場合」等の予め選択されたいくつかの条件項目にそれぞれ対応させて、それぞれ独立した行動モデルが設けられている。

【０１２３】そして、これら行動モデルは、それぞれ入力セマンティクスコンバータモジュール５９から認識結果が与えられたときや、最後の認識結果が与えられてから一定時間が経過したとき等に、必要に応じて後述のように感情モデル７３に保持されている対応する情動のパラメータ値や、本能モデル７４に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切換えモジュール７１に出力する。

【０１２４】なお、この実施の形態の場合、各行動モデルは、次の行動を決定する手法として、図１４に示すよ

うな１つのノード（状態）ＮＯＤＥ$_0$～ＮＯＤＥ$_n$から他のどのノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$に遷移するかを各ノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$に間を接続するアークＡＲＣ$_1$～ＡＲＣ$_{n_1}$に対してそれぞれ設定された遷移確率Ｐ$_1$～Ｐ$_n$に基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

【０１２５】具体的に、各行動モデルは、それぞれ自己の行動モデルを形成するノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$にそれぞれ対応させて、これらノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$毎に図１５に示すような状態遷移表８０を有している。

【０１２６】この状態遷移表８０では、そのノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$において遷移条件とする入力イベント（認識結果）が「入力イベント名」の行に優先順に列記され、その遷移条件についての更なる条件が「データ名」及び「データ範囲」の行における対応する列に記述されている。

【０１２７】したがって、図１５の状態遷移表８０で表されるノードＮＯＤＥ$_{100}$では、「ボールを検出（ＢＡＬＬ）」という認識結果が与えられた場合に、当該認識結果とともに与えられるそのボールの「大きさ（ＳＩＺＥ）」が「0から1000」の範囲であることや、「障害物を検出（ＯＢＳＴＡＣＬＥ）」という認識結果が与えられた場合に、当該認識結果とともに与えられるその障害物までの「距離（ＤＩＳＴＡＮＣＥ）」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

【０１２８】また、このノードＮＯＤＥ$_{100}$では、認識結果の入力がない場合においても、行動モデルが周期的に参照する感情モデル７３及び本能モデル７４にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル７３に保持された「喜び（Joy）」、「驚き（Surprise）」若しくは「悲しみ（Sadness）」の何れかのパラメータ値が「50から100」の範囲であるときには他のノードに遷移することができるようになっている。

【０１２９】また、状態遷移表８０では、「他のノードへの遷移確率」の欄における「遷移先ノード」の列にそのノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$から遷移できるノード名が列記されているとともに、「入力イベント名」、「データ名」及び「データの範囲」の行に記述された全ての条件が揃ったときに遷移できるほかの各ノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$への遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$に遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は１００［%］となっている。

【０１３０】したがって、図１５の状態遷移表８０で表

されるノードＮＯＤＥ１００では、例えば「ボールを検出（ＢＡＬＬ）」し、そのボールの「ＳＩＺＥ（大きさ）」が「0から1000」の範囲であるという認識結果が与えられた場合には、「30［％］」の確率で「ノードＮＯＤＥ１２０（node 120）」に遷移でき、そのとき「ＡＣＴＩＯＮ１」の行動が出力されることとなる。

【０１３１】各行動モデルは、それぞれこのような状態遷移表８０として記述されたノードＮＯＤＥ０～ ＮＯＤＥｎが幾つも繋がるようにして構成されており、入力セマンティクスコンバータモジュール５９から認識結果が与えられたとき等に、対応するノードＮＯＤＥ０～ＮＯＤＥｎの状態遷移表を利用して確率的に次の行動を決定し、決定結果を行動切換えモジュール７１に出力するようになされている。

【０１３２】図１２に示す行動切換えモジュール７１は、行動モデルライブラリ７０の各行動モデルからそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデルから出力された行動を選択し、当該行動を実行すべき旨のコマンド（以下、これを行動コマンドという。）をミドル・ウェア・レイヤ４０の出力セマンティクスコンバータモジュール６８に送出する。なお、この実施の形態においては、図１３において下側に表記された行動モデルほど優先順位が高く設定されている。

【０１３３】また、行動切換えモジュール７１は、行動完了後に出力セマンティクスコンバータモジュール６８から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール７２、感情モデル７３及び本能モデル７４に通知する。

【０１３４】一方、学習モジュール７２は、入力セマンティクスコンバータモジュール５９から与えられる認識結果のうち、「叩かれた」や「撫でられた」等、使用者からの働きかけとして受けた教示の認識結果を入力す＊る。

【０１３５】そして、学習モジュール７２は、この認識結果及び行動切換えモジュール７１からの通知に基づいて、「叩かれた（叱られた）」ときにはその行動の発現確率を低下させ、「撫でられた（誉められた）」ときにはその行動の発現確率を上昇させるように、行動モデルライブラリ７０における対応する行動モデルの対応する遷移確率を変更する。

【０１３６】他方、感情モデル７３は、「喜び（Joy）」、「悲しみ（Sadness）」、「怒り（Anger）」、「驚き（Surprise）」、「嫌悪（Disgust）」及び「恐れ（Fear）」の合計６つの情動について、各情動毎にその情動の強さを表すパラメータを保持している。そして、感情モデル７３は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール５９から与えられる「叩かれた」及び「撫でられた」等の特定の認識結果と、経過時間及び行動切換えモジュール７１からの通知と等に基づいて周期的に更新する。

【０１３７】具体的には、感情モデル７３は、入力セマンティクスコンバータモジュール５９から与えられる認識結果と、そのときのロボット装置１の行動と、前回更新してからの経過時間と等に基づいて所定の演算式により算出されるそのときのその情動の変動量を△Ｅ［ｔ］、現在のその情動のパラメータ値をＥ［ｔ］、その情動の感度を表す係数をｋｅとして、（１）式によって次の周期におけるその情動のパラメータ値Ｅ［ｔ＋１］を算出し、これを現在のその情動のパラメータ値Ｅ［ｔ］と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル７３は、これと同様にして全ての情動のパラメータ値を更新する。

【０１３８】

【数１】

$$E[t+1] = E[t] + ke \times \Delta E[t] \quad \cdots (1)$$

【０１３９】なお、各認識結果や出力セマンティクスコンバータモジュール６８からの通知が各情動のパラメータ値の変動量△Ｅ［ｔ］にどの程度の影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量△Ｅ［ｔ］に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量△Ｅ［ｔ］に大きな影響を与えるようになっている。

【０１４０】ここで、出力セマンティクスコンバータモジュール６８からの通知とは、いわゆる行動のフィードバック情報（行動完了情報）であり、行動の出現結果の情報であり、感情モデル７３は、このような情報によっても感情を変化させる。これは、例えば、「吠える」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール６８からの通知は、上述した学習モジュール７２にも入力されており、学習モジュール７２は、その通知に基づいて行動モデルの対応する遷移確率を変更する。

【０１４１】なお、行動結果のフィードバックは、行動切換えモジュレータ７１の出力（感情が付加された行動）によりなされるものであってもよい。

【０１４２】一方、本能モデル７４は、「運動欲（exercise）」、「愛情欲（affection）」、「食欲（appetite）」及び「好奇心（curiosity）」の互いに独立した４つの欲求について、これら欲求毎にその欲求の強さを表すパラメータを保持している。そして、本能モデル７４は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール５９から与えられる認識結果や、経過時間及び行動切換えモジュール７１からの通知等に基づいて周期的に更新する。

【０１４３】具体的には、本能モデル７４は、「運動

27

欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール６８からの通知等に基づいて所定の演算式により算出されるそのときのその欲求の変動量をΔＩ［ｋ］、現在のその欲求のパラメータ値をＩ［ｋ］、その欲求の感度を表す係数ｋ１として、所定周期で（２）式を用いて次の周期におけるその欲求のパラメータ値Ｉ［ｋ＋１］＊

28

＊を算出し、この演算結果を現在のその欲求のパラメータ値Ｉ［ｋ］と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル７４は、これと同様にして「食欲」を除く各欲求のパラメータ値を更新する。

【０１４４】

【数２】

$$I[k+1] = I[k] + ki \times \Delta I[k] \quad \cdots (2)$$

【０１４５】なお、認識結果及び出力セマンティクスコンバータモジュール６８からの通知等が各欲求のパラメータ値の変動量△Ｉ［ｋ］にどの程度の影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール６８からの通知は、「疲れ」のパラメータ値の変動量△Ｉ［ｋ］に大きな影響を与えるようになっている。

【０１４６】なお、本実施の形態においては、各情動及び各欲求（本能）のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また係数ｋｅ、ｋ１の値も各情動及び各欲求毎に個別に設定されている。

【０１４７】一方、ミドル・ウェア・レイヤ４０の出力セマンティクスコンバータモジュール６８は、図１１に示すように、上述のようにしてアプリケーション・レイヤ４１の行動切換えモジュール７１から与えられる「前進」、「喜ぶ」、「鳴く」又は「トラッキング（ボールを追いかける）」といった抽象的な行動コマンドを出力系６９の対応する信号処理モジュール６１～６７に与える。

【０１４８】そしてこれら信号処理モジュール６１～６７は、行動コマンドが与えられると当該行動コマンドに基づいて、その行動をするために対応するアクチュエータ２５１～２５ｎ（図２）に与えるべきサーボ指令値や、スピーカ２４（図２）から出力する音の音声データ及び又は「目」のＬＥＤに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト３２のバーチャル・ロボット３３及び信号処理回路１４（図２）を順次介して対応するアクチュエータ２５１～２５ｎ又はスピーカ２４又はＬＥＤに順次送出する。

【０１４９】このようにしてロボット装置１は、制御プログラムに基づいて、自己（内部）及び周囲（外部）の状況や、使用者からの指示及び働きかけに応じた自律的な行動ができる。したがって、上述した文字認識処理を実行するためプログラムを備えていないロボット装置に対しても、文字認識処理によって画像から抽出した文字の発音のしかたを音声認識処理によって周囲の音から認識された音声に基づいて決定する処理を実行するための制御プログラムを読み込ませることによって、図７に示した文字認識処理を実行させることができる。

【０１５０】このような制御プログラムは、ロボット装置が読取可能な形式で記録された記録媒体を介して提供される。制御プログラムを記録する記録媒体としては、磁気読取方式の記録媒体（例えば、磁気テープ、フロッピー（登録商標）ディスク、磁気カード）、光学読取方式の記録媒体（例えば、ＣＤ－ＲＯＭ、ＭＯ、ＣＤ－Ｒ、ＤＶＤ）等が考えられる。記録媒体には、半導体メモリ（いわゆるメモリカード（矩形型、正方形型等形状は問わない。）、ＩＣカード）等の記憶媒体も含まれる。また、制御プログラムは、いわゆるインターネット等を介して提供されてもよい。

【０１５１】これらの制御プログラムは、専用の読込ドライバ装置、又はパーソナルコンピュータ等を介して再生され、有線又は無線接続によってロボット装置１に伝送されて読み込まれる。また、ロボット装置は、半導体メモリ、又はＩＣカード等の小型化された記憶媒体のドライブ装置を備える場合、これら記憶媒体から制御プログラムを直接読み込むこともできる。ロボット装置１では、メモリカード２８から読み込むことができる。

【０１５２】なお、本発明は、上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。本実施の形態では、４足歩行のロボット装置に関して説明したが、ロボット装置は、２足歩行であってもよく、更に、移動手段は、脚式移動方式に限定されない。

【０１５３】以下に、本発明の別の実施の形態として示す人間型ロボット装置の詳細について説明する。図１６及び図１７には、人間型ロボット装置２００を前方及び後方の各々から眺望した様子を示している。更に、図１８には、この人間型ロボット装置２００が具備する関節自由度構成を模式的に示している。

【０１５４】図１６に示すように、人間型ロボット装置２００は、２本の腕部と頭部２０１を含む上肢と、移動動作を実現する２本の脚部からなる下肢と、上肢と下肢とを連結する体幹部とで構成される。

【０１５５】頭部２０１を支持する首関節は、首関節ヨー軸２０２と、首関節ピッチ軸２０３と、首関節ロール軸２０４という３自由度を有している。

【０１５６】また、各腕節は、肩関節ピッチ軸２０８と、肩関節ロール軸２０９と、上腕ヨー軸２１０と、肘関節ピッチ軸２１１と、前腕ヨー軸２１２と、手首関節ピッチ軸２１３と、手首関節ロール輪２１４と、手部２１５とで構成される。手部２１５は、実際には、複数本の指を含む多関節・多自由度構造体である。ただし、手部２１５の動作は人間型ロボット装置２００の姿勢制御や歩行制御に対する寄与や影響が少ないので、本明細書ではゼロ自由度と仮定する。したがって、各腕部は７自由度を有するとする。

【０１５７】また、体幹部は、体幹ピッチ軸２０５と、体幹ロール軸２０６と、体幹ヨー軸２０７という３自由度を有する。

【０１５８】また、下肢を構成する各々の脚部は、股関節ヨー軸２１６と、股関節ピッチ軸２１７と、股関節ロール軸２１８と、膝関節ピッチ軸２１９と、足首関節ピッチ軸２２０と、足首関節ロール軸２２１と、足部２２２とで構成される。本明細書中では、股関節ピッチ軸２１７と股関節ロール軸２１８の交点は、人間型ロボット装置２００の股関節位置を定義する。人体の足部２２２は、実際には多関節・多自由度の足底を含んだ構造体であるが、人間型ロボット装置２００の足底は、ゼロ自由度とする。したがって、各脚部は、６自由度で構成される。

【０１５９】以上を総括すれば、人間型ロボット装置２００全体としては、合計で３＋７×２＋３＋６×２＝３２自由度を有することになる。ただし、エンターテインメント向けの人間型ロボット装置２００が必ずしも３２自由度に限定される訳ではない。設計・制作上の制約条件や要求仕様等に応じて、自由度すなわち関節数を適宜増減することができることはいうまでもない。

【０１６０】上述したような人間型ロボット装置２００がもつ各自由度は、実際にはアクチュエータを用いて実装される。外観上で余分な膨らみを排してヒトの自然体形状に近似させること、２足歩行という不安定構造体に対して姿勢制御を行うことなどの要請から、アクチュエータは小型且つ軽量であることが好ましい。

【０１６１】図１９には、人間型ロボット装置２００の制御システム構成を模式的に示している。同図に示すように、人間型ロボット装置２００は、ヒトの四肢を表現した各機構ユニット２３０，２４０，２５０Ｒ／Ｌ，２６０Ｒ／Ｌと、各機構ユニット間の協調動作を実現するための適応制御を行う制御ユニット２８０とで構成される（ただし、Ｒ及びＬの各々は、右及び左の各々を示す接尾辞である。以下同様）。

【０１６２】人間型ロボット装置２００全体の動作は、制御ユニット２８０によって統括的に制御される。制御ユニット２８０は、ＣＰＵ（Central Processing Unit）やメモリ等の主要回路コンポーネント（図示しない）で構成される主制御部２８１と、電源回路や人間型ロボット装置２００の各構成要素とのデータやコマンドの授受を行うインターフェイス（何れも図示しない）などを含んだ周辺回路２８２とで構成される。この制御ユニット２８０の設置場所は、特に限定されない。図１９では体幹部ユニット２４０に搭載されているが、頭部ユニット２３０に搭載してもよい。或いは、人間型ロボット装置２００外に制御ユニット２８０を配備して、人間型ロボット装置２００の機体とは有線若しくは無線で交信するようにしてもよい。

【０１６３】図１９に示した人間型ロボット装置２００内の各関節自由度は、それぞれに対応するアクチュエータによって実現される。すなわち、頭部ユニット２３０には、首関節ヨー軸２０２、首関節ピッチ２０３、首関節ロール軸２０４の各々を表現する首関節ヨー軸アクチュエータA_2、首関節ピッチ軸アクチュエータA_3、首関節ロール軸アクチュエータA_4が配設されている。

【０１６４】また、頭部ユニット２３０には、外部の状況を撮像するためのＣＣＤ（ChargeCoupled Device）カメラが設けられているほか、前方に位置する物体までの距離を測定するための距離センサ、外部音を集音するためのマイク、音声を出力するためのスピーカ、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ等が配設されている。

【０１６５】また、体幹部ユニット２４０には、体幹ピッチ軸２０５、体幹ロール軸２０６、体幹ヨー軸２０７の各々を表現する体幹ピッチ軸アクチュエータA_5、体幹ロール軸アクチュエータA_6、体幹ヨー軸アクチュエータA_7が配設されている。また、体幹部ユニット２４０には、この人間型ロボット装置２００の起動電源となるバッテリを備えている。このバッテリは、充放電可能な電池によって構成されている。

【０１６６】また、腕部ユニット２５０Ｒ／Ｌは、上腕ユニット２５１Ｒ／Ｌと、肘関節ユニット２５２Ｒ／Ｌと、前腕ユニット２５３Ｒ／Ｌに細分化されるが、肩関節ピッチ軸8、肩関節ロール軸２０９、上腕ヨー軸２１０、肘関節ピッチ軸２１１、前腕ヨー軸２１２、手首関節ピッチ軸２１３、手首関節ロール軸２１４の各々表現する肩関節ピッチ軸アクチュエータA_8、肩関節ロール軸アクチュエータA_9、上腕ヨー軸アクチュエータA_{10}、肘関節ピッチ軸アクチュエータA_{11}、肘関節ロール軸アクチュエータA_{12}、手首関節ピッチ軸アクチュエータA_{13}、手首関節ロール軸アクチュエータA_{14}が配備されている。

【０１６７】また、脚部ユニット２６０Ｒ／Ｌは、大腿部ユニット２６１Ｒ／Ｌと、膝ユニット２６２Ｒ／Ｌと、脛部ユニット２６３Ｒ／Ｌに細分化されるが、股関節ヨー軸２１６、股関節ピッチ軸２１７、股関節ロール軸２１８、膝関節ピッチ軸２１９、足首関節ピッチ軸２２０、足首関節ロール軸２２１の各々を表現する股関節

31

ヨー軸アクチュエータＡ１６、股関節ピッチ軸アクチュエータＡ１７、股関節ロール軸アクチュエータＡ１８、膝関節ピッチ軸アクチュエータＡ１９、足首関節ピッチ軸アクチュエータＡ２０、足首関節ロール軸アクチュエータＡ２１が配備されている。各関節に用いられるアクチュエータＡ２，Ａ３・・・は、より好ましくは、ギア直結型で旦つサーボ制御系をワンチップ化してモータ・ユニット内に搭載したタイプの小型ＡＣサーボ・アクチュエータで構成することができる。

【０１６８】頭部ユニット２３０、体幹部ユニット２４０、腕部ユニット２５０、各脚部ユニット２６０などの各機構ユニット毎に、アクチュエータ駆動制御部の副制御部２３５，２４５，２５５Ｒ／Ｌ，２６５Ｒ／Ｌが配備されている。更に、各脚部２６０Ｒ，Ｌの足底が着床したか否かを検出する接地確認センサ２９１及び２９２を装着するとともに、体幹部ユニット２４０内には、姿勢を計測する姿勢センサ２９３を装備している。

【０１６９】接地確認センサ２９１及び２９２は、例えば足底に設置された近接センサ又はマイクロ・スイッチなどで構成される。また、姿勢センサ２９３は、例えば、加速度センサとジャイロ・センサの組み合わせによって構成される。

【０１７０】接地確認センサ２９１及び２９２の出力によって、歩行・走行などの動作期間中において、左右の各脚部が現在立脚又は遊脚何れの状態であるかを判別することができる。また、姿勢センサ２９３の出力により、体幹部分の傾きや姿勢を検出することができる。

【０１７１】主制御部２８１は、各センサ２９１～２９３の出力に応答して制御目標をダイナミックに補正することができる。より具体的には、副制御部２３５，２４５，２５５Ｒ／Ｌ，２６５Ｒ／Ｌの各々に対して適応的な制御を行い、人間型ロボット装置２００の上肢、体幹、及び下肢が協調して駆動する全身運動パターンを実現できる。

【０１７２】人間型ロボット装置２００の機体上での全身運動は、足部運動、ＺＭＰ（ZeroMoment Point）軌道、体幹運動、上肢運動、腰部高さなどを設定するとともに、これらの設定内容にしたがった動作を指示するコマンドを各副制御部２３５，２４５，２５５Ｒ／Ｌ，２６５Ｒ／Ｌに転送する。そして、各々の副制御部２３５，２４５，・・・等では、主制御部２８１からの受信コマンドを解釈して、各アクチュエータＡ２，Ａ３・・・等に対して駆動制御信号を出力する。ここでいう「ＺＭＰ」とは、歩行中の床反力によるモーメントがゼロとなる床面上の点のことであり、また、「ＺＭＰ軌道」とは、例えば人間型ロボット装置２００の歩行動作期間中にＺＭＰが動く軌跡を意味する。

【０１７３】歩行時には、重力と歩行運動に伴って生じる加速度によって、歩行系から路面には重力と慣性力、並びにこれらのモーメントが作用する。いわゆる「ダランベールの原理」によると、それらは路面から歩行系への反作用としての床反力、床反力モーメントとバランスする。力学的推論の帰結として、足底接地点と路面の形成する支持多角形の辺上或いはその内側にピッチ及びロール軸モーメントがゼロとなる点、すなわち「ＺＭＰ（Zero Moment Point）」が存在する。

32

【０１７４】脚式移動ロボットの姿勢安定制御や歩行時の転倒防止に関する提案の多くは、このＺＭＰを歩行の安定度判別の規範として用いたものである。ＺＭＰ規範に基づく２足歩行パターン生成は、足底着地点を予め設定することができ、路面形状に応じた足先の運動学的拘束条件を考慮しやすいなどの利点がある。また、ＺＭＰを安定度判別規範とすることは、力ではなく軌道を運動制御上の目標値として扱うことを意味するので、技術的に実現可能性が高まる。なお、ＺＭＰの概念並びにＺＭＰを歩行ロボットの安定度判別規範に適用する点については、Miomir Vukobratovic著“LEGGED LOCOMOTION ROBOTS”（加藤一郎外著『歩行ロボットと人工の足』（日刊工業新聞社））に記載されている。

【０１７５】一般には、４足歩行よりもヒューマノイドのような２足歩行のロボットの方が、重心位置が高く、且つ、歩行時のＺＭＰ安定領域が狭い。したがって、このような路面状態の変化に伴う姿勢変動の問題は、２足歩行ロボットにおいてとりわけ重要となる。

【０１７６】以上のように、人間型ロボット装置２００は、各々の副制御部２３５，２４５，・・・等が、主制御部２８１からの受信コマンドを解釈して、各アクチュエータＡ２，Ａ３・・・に対して駆動制御信号を出力し、各ユニットの駆動を制御している。これにより、人間型ロボット装置２００は、目標の姿勢に安定して遷移し、安定した姿勢で歩行できる。

【０１７７】また、人間型ロボット装置２００における制御ユニット２８０では、上述したような姿勢制御のほかに、加速度センサ、タッチセンサ、接地確認センサ等の各種センサ、及びＣＣＤカメラからの画像情報、マイクからの音声情報等を統括して処理している。制御ユニット２８０では、図示しないが加速度センサ、ジャイロ・センサ、タッチセンサ、距離センサ、マイク、スピーカなどの各種センサ、各アクチュエータ、ＣＣＤカメラ及びバッテリが各々対応するハブを介して主制御部２８１と接続されている。

【０１７８】主制御部２８１は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部インターフェイスを介してＤＲＡＭ内の所定位置に順次格納する。また、主制御部２８１は、バッテリから供給されるバッテリ残量を表すバッテリ残量データを順次取り込み、これをＤＲＡＭ内の所定位置に格納する。ＤＲＡＭに格納された各センサデータ、画像データ、音声データ及びバッテリ残量データは、主制御部２８１がこの人間型ロボット装置

２００の動作制御を行う際に利用される。

【０１７９】主制御部２８１は、人間型ロボット装置２００の電源が投入された初期時、制御プログラムを読み出し、これをＤＲＡＭに格納する。また、主制御部２８１は、上述のように主制御部２８１よりＤＲＡＭに順次格納される各センサデータ、画像データ、音声データ及びバッテリ残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無などを判断する。更に、主制御部２８１は、この判断結果及びＤＲＡＭに格納した制御プログラムに基づいて自己の状況に応じて行動を決定するとともに、当該決定結果に基づいて必要なアクチュエータを駆動させることにより人間型ロボット装置２００に、いわゆる「身振り」、「手振り」といった行動をとらせる。

【０１８０】したがって、人間型ロボット装置２００は、制御プログラムに基づいて自己及び周囲の状況を判断し、使用者からの指示及び働きかけに応じて自律的に行動できる。また、人間型ロボット装置２００は、ＣＣＤカメラにおいて撮像された画像から抽出した文字の発音のしかた（読み方）を、抽出された文字から推定される読み方と集音マイクにおいて集音された音声とをマッチングして決定する。したがって、人間型ロボット装置２００の音声認識の精度が向上し、新規単語が音声認識用辞書に登録できる。

【０１８１】

【発明の効果】以上詳細に説明したように、本発明に係るロボット装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【０１８２】本発明に係るロボット装置は、撮像手段において撮像された画像から抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与した複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較して最も近い音声波形を抽出した文字の発音のしかたであるとして決定する。

【０１８３】したがって、本発明に係るロボット装置によれば、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１８４】また、本発明に係るロボット装置は、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

【０１８５】したがって、本発明に係るロボット装置によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合に属性情報が入力できなかったという不都合が改善される。

【０１８６】また、本発明に係る文字認識装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【０１８７】したがって、本発明に係る文字認識装置によれば、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１８８】また、本発明に係る文字認識装置は、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備

え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

【０１８９】したがって、本発明に係る文字認識装置によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

【０１９０】また、本発明に係る文字認識方法は、被写体を撮像する撮像工程と、撮像工程において撮像された画像から所定パターンの画像を抽出する画像認識工程と、周囲の音を取得する集音工程と、集音工程において取得された音から音声を認識する音声認識工程と、画像認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備える。

【０１９１】したがって、本発明に係る文字認識方法によれば、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１９２】また、本発明に係る文字認識方法によれば、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

【０１９３】したがって、本発明に係る文字認識方法によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

【０１９４】更に、本発明に係る制御プログラムは、被写体を撮像する撮像処理と、撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、周囲の音を取得する集音処理と、集音処理によって取得された音から音声を認識する音声認識処理と、画像認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、発音情報生成処理によって生成された各音声波形と音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させる。

【０１９５】したがって、本発明に係る制御プログラムによれば、ロボット装置は、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上される。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１９６】また、上述の制御プログラムを記録媒体に記録して提供することによって、この記録媒体を読込可能で画像認識手段と音声認識手段とを備える音声認識装置としての機能を有する電子機器に対して、新規単語を登録する際の認識精度が向上される。これにより、正確な発音のしかたが記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【図面の簡単な説明】

【図１】本発明の一構成例として示すロボット装置の外観を示す外観図である。

【図２】本発明の一構成例として示すロボット装置の構成を示す構成図である。

【図３】本発明の一構成例として示すロボット装置における画像音声認識部の構成を示す構成図である。

【図４】本発明の一構成例として示すロボット装置の音声認識用辞書を説明する図である。

【図５】本発明の一構成例として示すロボット装置の単語読み属性テーブルを説明する図である。

【図６】本発明の一構成例として示すロボット装置の文字読みテーブルを説明する図である。

【図７】本発明の一構成例として示すロボット装置が新規単語を音声認識用辞書に登録する処理を説明するフローチャートである。

【図８】本発明の一構成例として示すロボット装置の新規単語用認識用辞書を説明する図である。

【図９】本発明の一構成例として示すロボット装置が認識した文字列の発音のしかた（読み方）を生成する処理を説明するフローチャートである。

【図１０】本発明の一構成例として示すロボット装置の制御プログラムのソフトウェア構成を示す構成図である。

37

【図１１】本発明の一構成例として示すロボット装置の制御プログラムのうち、ミドル・ウェア・レイヤの構成を示す構成図である。

【図１２】本発明の一構成例として示すロボット装置の制御プログラムのうち、アプリケーション・レイヤの構成を示す構成図である。

【図１３】本発明の一構成例として示すロボット装置の制御プログラムのうち、行動モデルライブラリの構成を示す構成図である。

【図１４】本発明の一構成例として示すロボット装置の行動を決定するためのアルゴリズムである有限確率オートマトンを説明する模式図である。

【図１５】本発明の一構成例として示すロボット装置の行動を決定するための状態遷移条件を表す図である。

【図１６】本発明の一構成例として示す人間型ロボット装置の前方からみた外観を説明する外観図である。

【図１７】本発明の一構成例として示す人間型ロボット装置の後方からみた外観を説明する外観図である。

【図１８】本発明の一構成例として示す人間型ロボット装置の自由度構成モデルを模式的に示す図である。

【図１９】本発明の一構成例として示す人間型ロボット装置の制御システム構成を説明する図である。

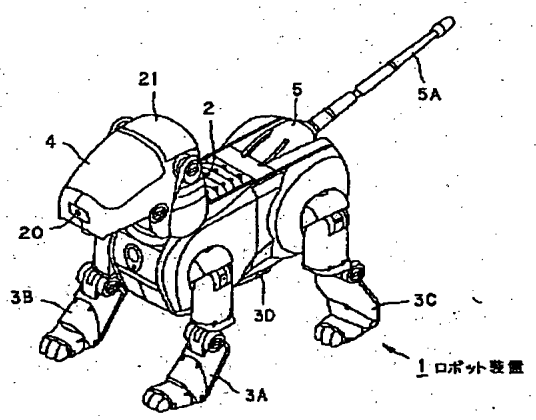
【図２０】図２０（ａ）は、「音素」を基本単位とするガーベージモデルを適用した従来の音声認識方法を示す模式図であり、図２０（ｂ）は、「かな」を基本単位とするガーベージモデルを適用した従来の音声認識方法を示す模式図である。

38

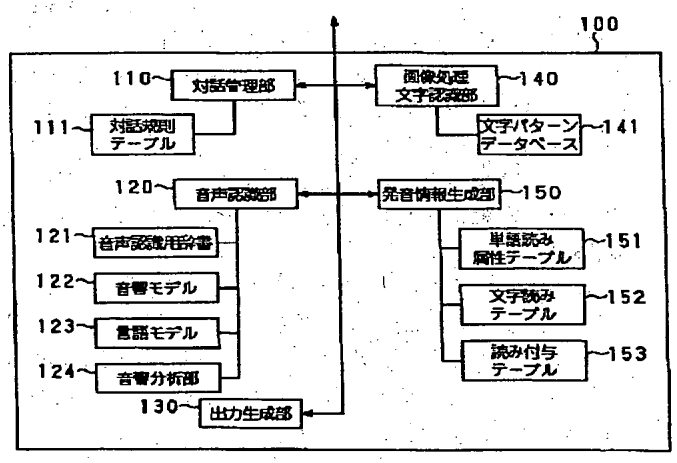
【符号の説明】

１　ロボット装置、２　胴体部ユニット、３Ａ，３Ｂ，３Ｃ，３Ｄ　脚部ユニット、４　頭部ユニット、５　尻尾部ユニット、１０　ＣＰＵ、１１　ＤＲＡＭ、１２　フラッシュＲＯＭ、１３　ＰＣカードインターフェイス回路、１４　信号処理回路、１５　内部バス、１６　コントロール部、１７　バッテリ、１８角速度センサ、１９　加速度センサ、２０　ＣＣＤカメラ、２１　タッチセンサ、２２　距離センサ、２３　マイク、２４　スピーカ、２５$_1$～２５$_n$　アクチュエータ、２６$_1$～２６$_n$　ポテンショメータ、２７$_1$～２７$_n$　ハブ、２８メモリカード、１００　画像音声認識部、１１０　対話管理部、１１１対話規則テーブル、１２０　音声認識部、１２１　音声認識用辞書、１２２　音響モデル、１２３　言語モデル、１２４　音響分析部、１３０　出力生成部、１４０　画像処理文字認識部、１４１　文字パターンデータベース、１５０　発音情報生成部、１５１　単語読み属性テーブル、１５２　文字読みテーブル、１５３　読み付与テーブル、２００　人間型ロボット装置

【図１】



【図３】



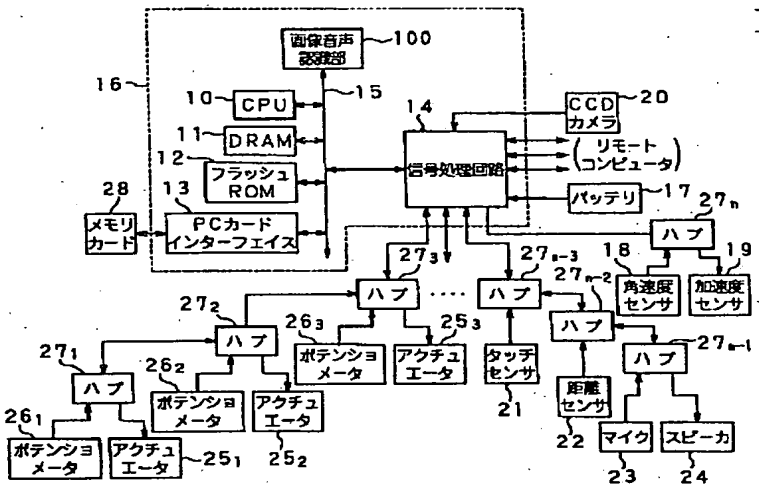
【図４】

単語シンボル	PLU列
五反田	gotaNda
電車	deNsha
に	ni
行き	Iki
たい	tai
・・・	・・・

【図５】

単語	読み	属性
品川	しながわ	（地名）
北品川	きたしながわ	（地名）
青物横丁	あおものよこちょう	（地名）
・・・	・・・	・・・
佐藤	さとう	（姓）
鈴木	すずき	（姓）
・・・	・・・	・・・
すべすべ饅頭	すべすべまんじゅうがに	（動物）
大福海鼠	おおいかりなまこ	（動物）
・・・	・・・	・・・

【図２】



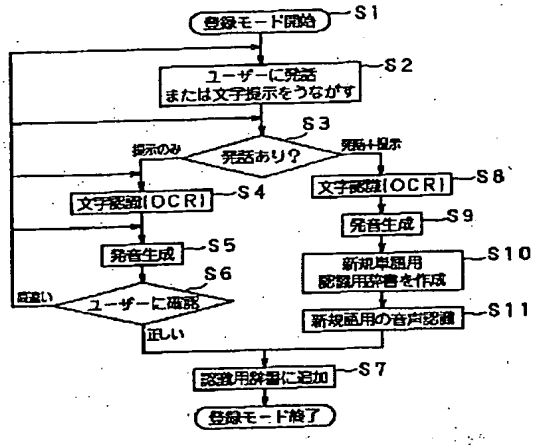
【図６】

単語	読み
北	きた,ホク
品	しな,ヒン,ホン
川	かわ,セン
比	くら,ヒ
品	あきら,ショウ
：	：
？	はてな,クエスチョン
@	アット
：	：
A	エー,ア
B	ビー,ブ
：	：

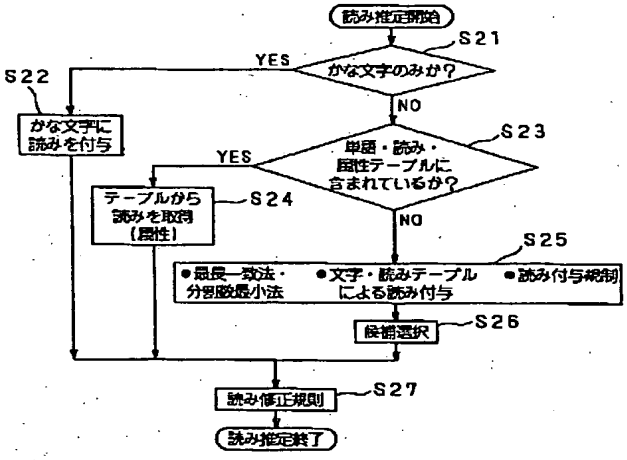
【図８】

単語シンボル	PLU列
北品川	k i t a s h i n a g a w a
北品川	h i a h o : k a w a
比品川	k u r a a k i r a a w a

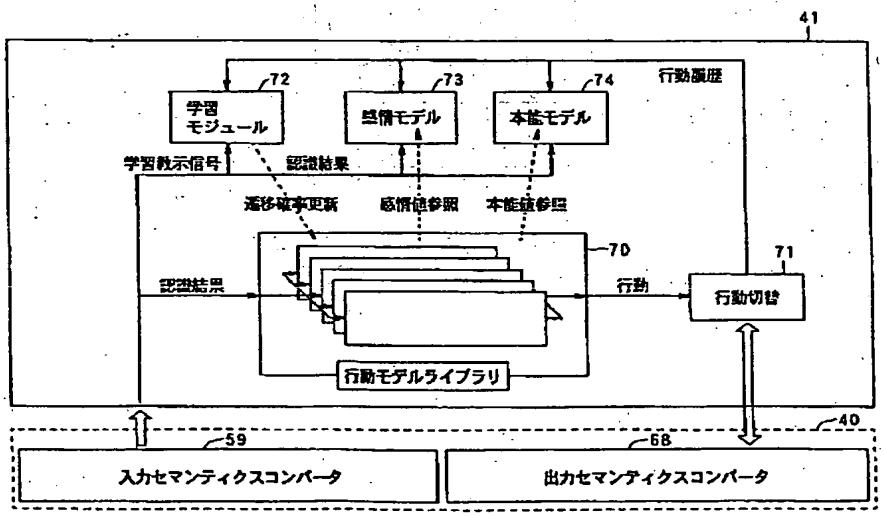
【図７】



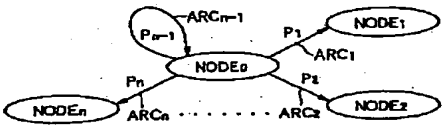
【図９】



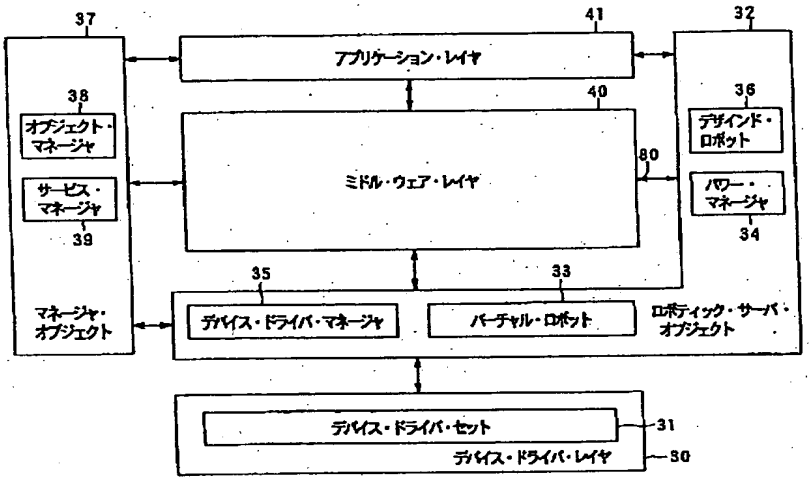
【図１２】



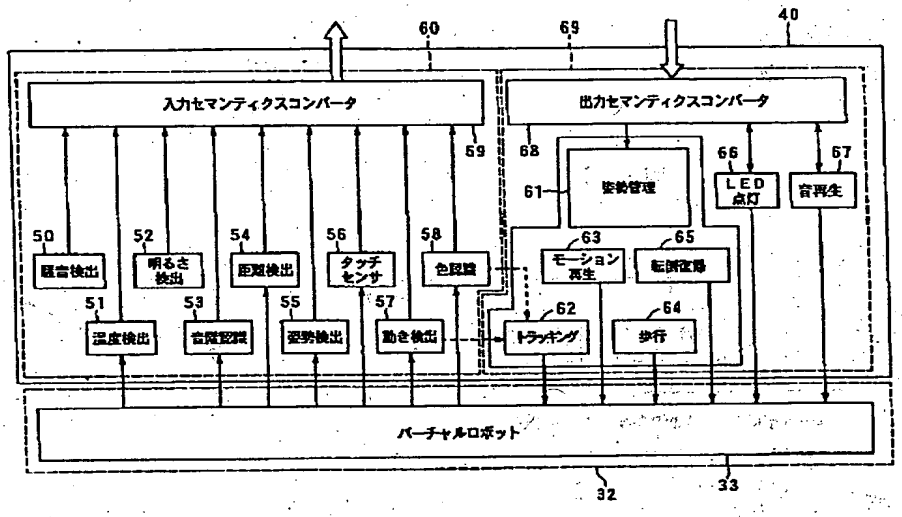
【図１４】



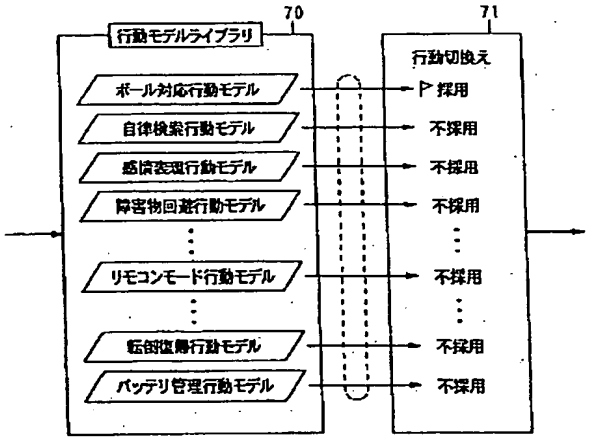
【図１０】


37
41
32
アプリケーション・レイヤ
38
オブジェクト・マネージャ
40
36
デザインド・ロボット
ミドル・ウェア・レイヤ
80
サービス・マネージャ
パワー・マネージャ
39
34
35
33
マネージャ・オブジェクト
デバイス・ドライバ・マネージャ
バーチャル・ロボット
ロボティック・サーバ・オブジェクト
デバイス・ドライバ・セット
31
デバイス・ドライバ・レイヤ
30

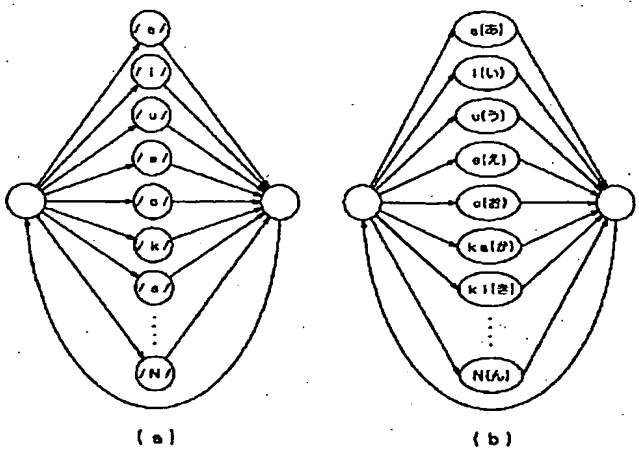
【図１１】


60
69
40
入力セマンティクスコンバータ
出力セマンティクスコンバータ
59
68
66
67
61
姿勢管理
LED点灯
音再生
50
52
54
56
58
63
65
発話検出
明るさ検出
距離検出
タッチセンサ
色認識
モーション再生
転倒復帰
51
53
55
57
62
64
温度検出
音階認識
姿勢検出
動き検出
トラッキング
歩行
バーチャルロボット
32
33

【図１３】



【図２０】

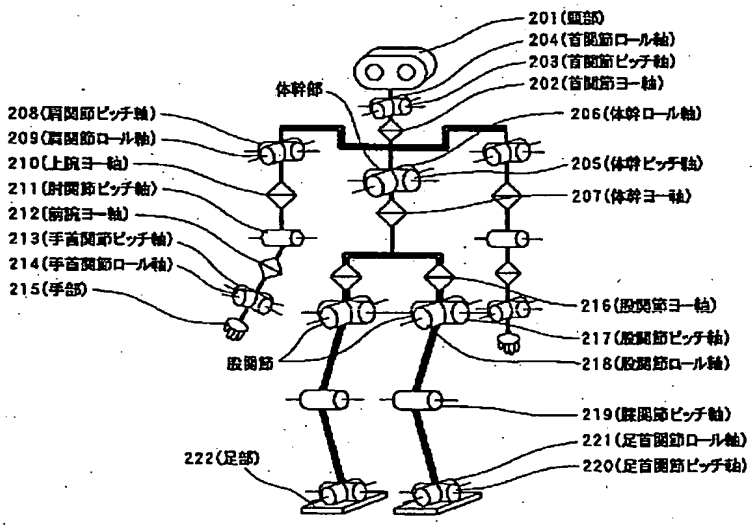


【図１５】

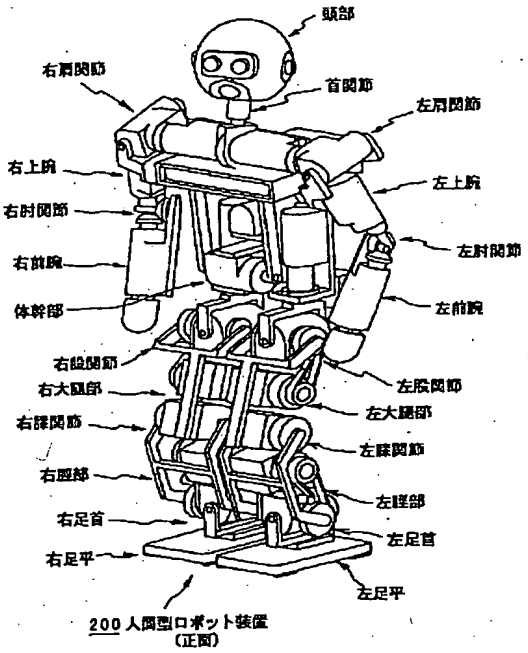
	入力イベント名	データ名	データの範囲	他のノードへの遷移確率 Di				
node 100				A	B	C	D	n
遷移先ノード				node 120	node 120	node 1000		node 600
出力行動				ACTION 1	ACTION 2	MOVE BACK		ACTION 4
1	BALL	SIZE	0.1000	30%				
2	PAT				40%			
3	HIT				20%			
4	MOTION					50%		
5	OBSTACLE	DISTANCE	0.100			100%		
6		JOY	50.100					
7		SURPRISE	50.100					
8		SADNESS	50.100					

80

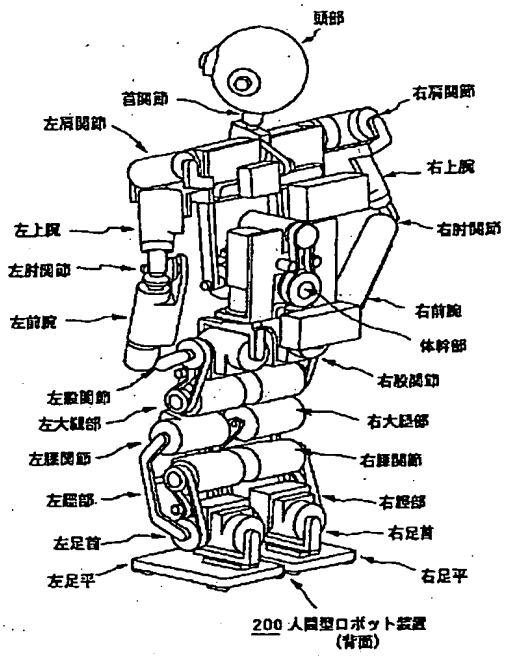
【図１８】



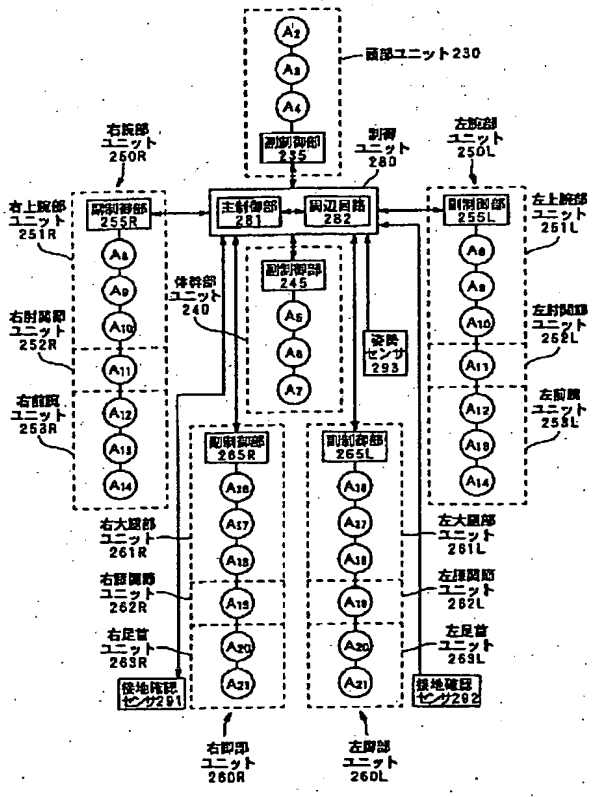
【図１６】


頭部
右肩関節
首関節
左肩関節
右上腕
左上腕
右肘関節
左肘関節
右前腕
左前腕
体幹部
右股関節
左股関節
右大腿部
左大腿部
右膝関節
左膝関節
右脛部
左脛部
右足首
左足首
右足平
左足平
200 人間型ロボット装置
(正面)

【図１７】


頭部
首関節
右肩関節
左肩関節
右上腕
左上腕
右肘関節
左肘関節
右前腕
左前腕
体幹部
右股関節
左股関節
右大腿部
左大腿部
右膝関節
左膝関節
右脛部
左脛部
右足首
左足首
右足平
左足平
200 人間型ロボット装置
(背面)

【図１９】


頭部ユニット230
副制御部 235
制御ユニット 280
右腕部ユニット 250R
左腕部ユニット 250L
制御部 255R
主制御部 281
周辺回路 282
制御部 255L
右上腕部ユニット 251R
左上腕部ユニット 251L
体幹部ユニット 240
副制御部 245
姿勢センサ 293
右肘関節ユニット 252R
左肘関節ユニット 252L
右前腕ユニット 253R
左前腕ユニット 253L
制御部 265R
制御部 265L
右大腿部ユニット 261R
左大腿部ユニット 261L
右膝関節ユニット 262R
左膝関節ユニット 262L
右足首ユニット 263R
左足首ユニット 263L
接地確認センサ291
接地確認センサ292
右脚部ユニット 260R
左脚部ユニット 260L

フロントページの続き

(51) Int. Cl.[7]	識別記号	F I	テーマコード（参考）
Ｇ１０Ｌ　15/22		Ｇ１０Ｌ　3/00	５２１Ｖ
15/24			５７１Ｑ
			５７１Ｔ
			５３１Ｑ

(72)発明者　河本　献太
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(72)発明者　佐部　浩太郎
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(72)発明者　大橋　武史
東京都品川区北品川６丁目７番35号　ソニー株式会社内

Ｆターム（参考）　5B064 AA07 FA16
5D015 GG03 HH23 KK02 KK04 LL07 LL11

This Page is inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning documents *will not* correct images problems checked, please do not report the problems to the IFW Image Problem Mailbox